On the robust automatic computation of speaking and articulation duration in ALS patients versus healthy controls

Jackson Liscombe[1], Alexander Ocampo[1], Hardik Kothare[1], Oliver Roesler[1], Michael Neumann[1], Andrew Cornish[1], Doug Habberstad[1], David Pautler[1], David Suendermann-Oeft[1], and Vikram Ramanarayanan[1,2]
[1]Modality.AI, [2]University of California, San Francisco
Contact: jackson.liscombe@modality.ai

Objective

To optimize articulation boundary detection for calculation of prosodic features.

Annotation

All speech and non-speech regions were marked using Praat. ***Speaking duration (SD)*** = time between start of first and end of last speech region. ***Articulation duration (AD)*** = sum of all speech regions. Inter-annotator agreement on 45 utterances from 4 sessions was 0.42s RMSE.

Automatic Detection Methods

Praat Sound: To TextGrid (silences)...

- min_pitch: 50–500Hz
- time_step: 0.0–5.0s
- silence_threshold: -(1–70)dB
- min_silent_interval: 0.1–1.0s
- min_sounding_interval: 0.1–1.0s

WebRTC VAD

- aggressiveness: 0–3
- frame_duration: 10–30ms
- padding_duration: 100–2000ms

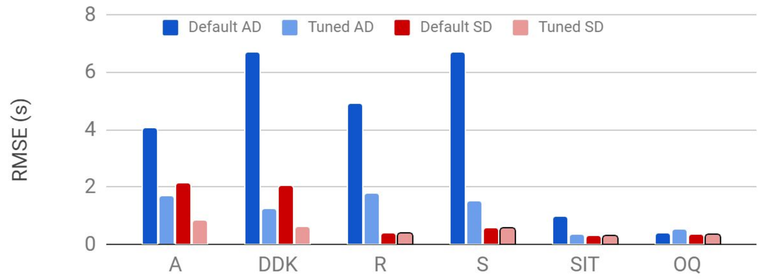
Task Descriptions

OQ	Have you had any challenges when speaking, salivating, or swallowing? If so, please briefly describe any difficulties.
A	Please take a deep breath and then say "aaa" until you run out of breath.
DDK	Please take a deep breath and say "pataka" over and over until you run out of breath.
SIT	Please say, "The job provides many benefits." [Repeated 5 more times with different sentences.]
R	Please read the text aloud to me, to the best of your ability. [Participant shown text of passage about bamboo.]
S	Please describe what you see happening in this picture. Please try to speak for at least one minute.

Corpus

2,231 turns from 195 sessions collected between July 28, 2020 and February 22, 2021 from users of NEMSI to assess symptoms of ALS.

Comparing Overall-Tuned Praat Per Task (AD and SD)



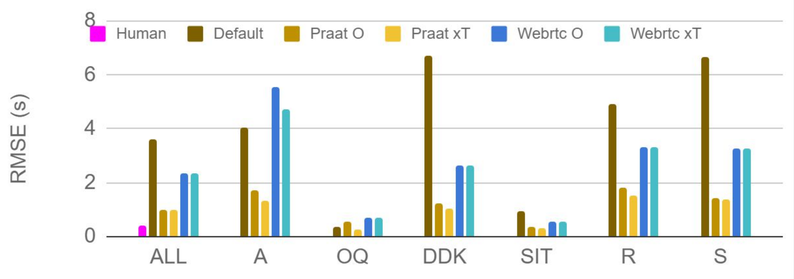
Results

The only difference in tuned parameters was silence threshold. For each, these default settings were best:

- min_pitch = 100Hz
- time_step = 0.0s
- Minimum silent interval (s): 0.1
- Minimum sounding interval (s): 0.1

For **speaking duration**, the default setting of -25dB was best but for **articulation duration**, a new setting of **-36dB silence threshold** was found to be optimal.

Comparing Task-Tuned Predictors (AD)



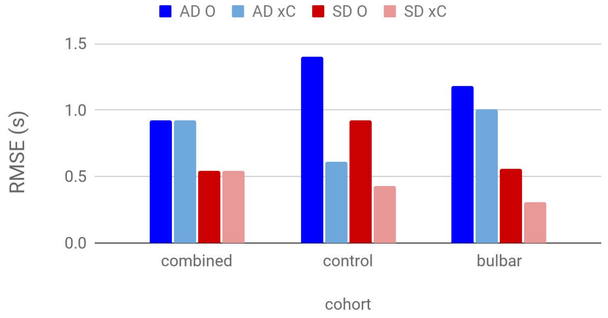
Results

Tuned Praat outperformed WebRTC VAD.

Tuning per task yielded slight improvement in performance.

Task	silence_threshold	min_silent_interval	min_sounding_interval
OQ	-29	0.1	0.1
A	-46	0.1	0.1
DDK	-37	0.1	0.1
SIT	-26	0.4	0.1
R	-35	0.05	0.03
S	-37	0.1	0.1

Comparing Cohort-Tuned Predictors (AD and SD)



Corpus and Results

Control cohort: not diagnosed with ALS, ALSFRS-R score of 48, 951 turns from 86 sessions.

Bulbar cohort: diagnosed with ALS, ALSFRS-R Bulbar subscore < 12, 526 turns from 48 sessions.

For **speaking duration**, performance was better for **bulbar** users with a slightly **lower silence threshold** compared to control users (-24db vs -26db) and a minimum sounding interval of 0.05s instead of 0.1s for the control cohort.

For **articulation duration**, performance was better for **bulbar** users with a **higher silence threshold** compared to control users (-38db vs -28db) and a **lower minimum silence interval** of 0.1s compared to 0.4s for control.